AI-powered clinical decision support system for monitoring patients with of age-related macular degeneration

María Pulido Agüero [1], Alberto J. Beltrán Carrero [1], Javier Torresano Rodríguez [2], Enrique Maldonado Belmonte [3], Esther Santos Vicente [2], Álvaro Caballero Sastre [1], María J. Ledesma-Carbayo[1,4], Juan-J. Gómez-Valverde [1,4]

[1] Biomedical Image Technologies (BIT), ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain, maria.pulidoa@upm.es

[2] Instituto Provincial de Oftalmología, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[3] Subdirección de Sistemas de Información , Hospital General Universitario Gregorio Marañón, Madrid, Spain

[4] Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain

Abstract

Age-related macular degeneration (AMD) is the main cause of legal blindness in elderly populations of developed countries. Its neovascular form, the most aggressive, requires frequent OCT monitoring and repeated anti-VEGF injections, creating high workload and variability in clinical decisions. We present STEP-AMD, an AI-powered clinical decision support system (CDSS) designed to assist in the follow-up of neovascular AMD patients. The system integrates OCT analysis with clinical data stored in a relational database and automatically generates structured PDF reports. These longitudinal reports summarize fluid biomarkers, visual acuity, treatment dates, and predicted therapy response, providing specialists with a comprehensive overview for personalized treatment planning. The system was developed and deployed at Hospital General Universitario Gregorio Marañón using a dataset of 503 OCT studies from 190 patients to train segmentation and prediction models. In a proof-of-concept study with two retina specialists evaluating 17 patients (21 eyes), STEP-AMD reduced average decision-making time by 39%, improved inter-rater agreement (Cohen's kappa from 0.42 to 0.71), and increased diagnostic accuracy on average from 0.42 to 0.56. Clinicians rated the AI-generated reports highly in usability and usefulness. These results highlight the potential of STEP-AMD to standardize follow-up, reduce workload, and improve efficiency in AMD management.

1. Introduction

Age-related macular degeneration (AMD) is one of the most common causes of severe visual impairment in industrialized countries. The neovascular subtype, although less frequent, is the most aggressive and accounts for the majority of vision loss cases [1]. Management relies on repeated intravitreal injections of anti-VEGF agents and continuous follow-up with optical coherence tomography (OCT), which places a considerable burden on healthcare systems and often leads to variability in clinical decisions. Disease activity is primarily assessed through OCT biomarkers, including intraretinal fluid (IRF), subretinal fluid (SRF), and pigment epithelial detachment (PED) [1, 2], which are central for diagnosis, monitoring, and guiding therapeutic interventions [3, 2].

Artificial intelligence (AI) has rapidly expanded in ophthalmology, with studies demonstrating strong performance in OCT biomarker segmentation and treatment outcome prediction for AMD patients [3, 4]. Recent work has also shown that deep learning models can integrate OCT data with clinical variables to predict visual outcomes after anti-VEGF therapy in neovascular AMD [5], and that machine learning approaches can estimate patient-specific treatment demand in treat-and-extend regimens [6]. Nevertheless, most of these methods remain primarily algorithmic, with limited validation in real-world workflows and without integration into decision support systems accessible to clinicians.

In parallel, a few commercial systems have obtained regulatory approval, including IDx-DR (Digital Diagnostics, USA) and EyeArt (Eyenuk, USA) for diabetic retinopathy detection, and more recently the Notal Vision Home OCT System (Notal Vision, USA), the only FDA-approved OCT-based AI for remote monitoring of AMD. In Europe, CE-marked solutions such as RetInSight Fluid Monitor (RetInSight, Austria) extend OCT-based analysis, but their routine clinical use remains limited. Despite these advances, few systems specifically target AMD management, and most tools are restricted to screening or isolated biomarker quantification, without integrating longitudinal clinical data into structured reports for daily practice.

In this work, we present STEP-AMD, a clinical decision support system (CDSS) that combines AI-based OCT analysis with clinical data to automatically generate structured PDF reports. These reports summarize biomarker quantification, visual acuity, treatment history, and predicted response, providing clinicians with a comprehensive overview for follow-up. A proof-of-concept evaluation with retina specialists demonstrates the potential of STEP-AMD to improve efficiency and consistency in AMD care.

2. Materials

2.1. Dataset

A real-world longitudinal dataset of radial OCT scans was collected at the Instituto Provincial de Oftalmología, Hospi-

ISBN: 978-84-09-80259-3

tal General Universitario Gregorio Marañón (HGUGM) using a Topcon 3D OCT-1 Maestro device. The cohort included patients diagnosed with neovascular AMD who initiated anti-VEGF therapy with bevacizumab between 2021 and 2023. For each patient, OCT scans were acquired immediately before treatment, after the initial loading phase of three monthly bevacizumab injections, and at the subsequent follow-up visit, together with clinical data such as visual acuity and treatment records. All examinations were conducted under HGUGM ethics approval, with written informed consent obtained from all participants.

In total, 503 OCT studies from 190 patients were collected. Each study included acquisitions from both eyes when available, resulting in 1006 individual examinations. Each examination consisted of 12 radial B-scans centered on the macula.

2.2. Pre-trained DL models

The system integrates three automatic algorithms already developed for this project with two main purposes: segmentation of fluid biomarkers and prediction of treatment response.

Segmentation model: A 2D MedNeXt-Base-kernel3 network [7] was trained to automatically segment IRF, SRF, and PED in OCT scans. An example result is shown in Figure 1. This model was developed with a total of 111 OCT studies from 90 patients, not included in the longitudinal dataset.

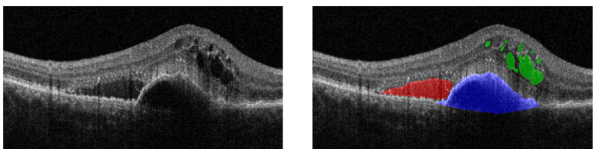

Figure 1: Example of a radial OCT B-scan from a neovascular AMD patient. Left: original image. Right: segmentation generated by the MedNeXt-Base-kernel3 model, highlighting intraretinal fluid (IRF, green), subretinal fluid (SRF, red), and pigment epithelial detachment (PED, blue).

Prediction models: Two Multiple Instance Learning (MIL) networks, *MultiPhase0* and *MultiPhase1*, were developed to classify patients as Good Responder (substantial fluid reduction), Sub-optimal Responder (partial reduction), or Non-responder (no reduction) to bevacizumab therapy [8]. The first model uses only the baseline OCT, while the second incorporates both baseline and post-loading scans, providing temporal information for prediction. Both models were trained and evaluated on the complete longitudinal dataset, comprising 503 OCT studies.

2.3. Relational database

A MySQL-based relational database, named *BitScreen Retina*, was used to integrate OCT data, results from the DL models, and clinical information such as visual acuity and treatment history. The database consists of 32 interrelated tables organized to manage users, patients, studies, examinations, and diagnostics. This structure links imaging and clinical data with AI-derived outputs such as fluid biomarker quantification and predicted treatment response, enabling longitudinal tracking of patient evolution and supporting the automatic generation of structured PDF reports.

2.4. Hardware and Software

The system was deployed on a virtual (VM) (Ubuntu 22.04 LTS, 8 CPUs, 16 GB RAM, 100 GB storage) provided by the HGUGM IT Department, which hosted the MySQL database, CDSS services, and inference environments. Training and heavy computation were carried out on the BIT-UPM high-performance cluster with GPU-enabled nodes, accessed through a secure VPN. The software was implemented in Python using Anaconda, with two Conda environments: one (Python 3.11.11) for the STEP-AMD workflow and prediction models, and another (Python 3.10.16) for the MedNeXt segmentation model. The overall system architecture is shown in Figure 2.

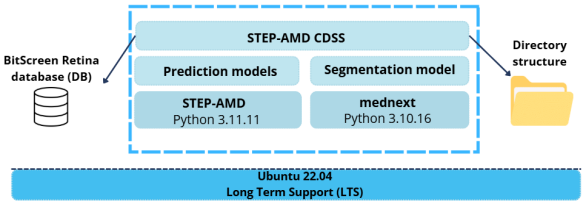

Figure 2: System architecture of STEP-AMD.

3. Methods

3.1. STEP-AMD CDSS

Figure 3 summarizes the workflow of the proposed CDSS. The system takes two inputs: clinical data collected during each visit and radial OCT scans from both eyes, and produces a single output: a structured PDF medical report for the ophthalmologist.

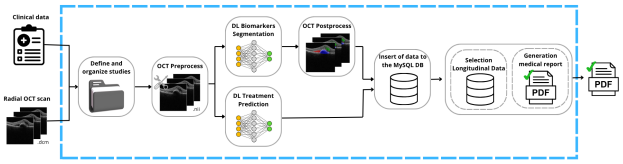

Figure 3: Workflow of the STEP-AMD system.

The system is modular, with each block implemented as a Python script executed sequentially by a main controller. First, incoming studies are organized and OCT scans are converted from DICOM to NIfTI format. Fluid biomarkers (IRF, SRF, PED) are segmented using the DL model, and the resulting masks are post-processed to compute biomarker areas. In parallel, OCT scans are analyzed with DL prediction models when they correspond to pre- or post-loading phases of anti-VEGF therapy. All outputs, together with clinical data, are stored in the *BitScreen Retina* database. Finally, a PDF report is generated by retrieving information from the last studies of each patient, providing a longitudinal overview

of disease progression.

3.2. Ophthalmologists' evaluation

Two retina specialists from HGUGM participated in the evaluation of the AI-generated reports to ensure clinical relevance and usability. A proof-of-concept study was conducted on 17 AMD patients from the collected longitudinal dataset (21 eyes; mean age 82.9 ± 8.1 years; balanced gender). Treatment decisions (continue, adjust, change, or stop therapy) were compared with and without AI assistance, recording decision accuracy, confidence, and time. Performance was measured against clinical practice as the gold standard using precision, recall, F1-score, accuracy, Cohen's kappa, and change ratio. In addition, a 23-item survey assessed usability, usefulness, confidence, ethical aspects, satisfaction, and potential adoption, providing insights into both technical performance and clinical acceptance.

4. Results and Discussion

4.1. AI-assisted PDF reports

The STEP-AMD system successfully integrates OCT data and clinical variables into a workflow that produces structured PDF reports for longitudinal follow-up. Each report (Figure 4) includes patient information, progression plots for visual acuity and biomarker areas (IRF, SRF, PED), injection dates, and treatment response predictions. In addition, segmentation analysis provides quantitative tables and representative OCT B-scans. A technical notes section documents the AI models used and reminds clinicians that the results are decision-support outputs. Reports are digitally signed and stored securely, ensuring traceability and compliance.

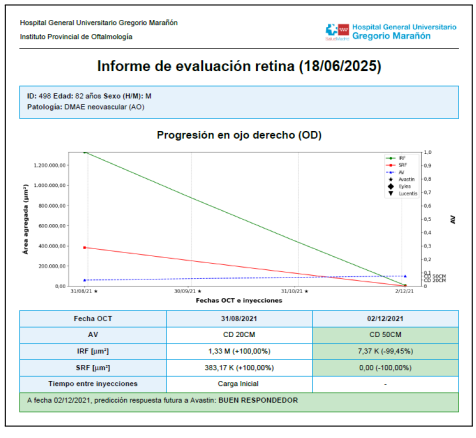

Figure 4: Example of the progression plots included in the PDF medical report generated by STEP-AMD, combining imaging biomarkers, clinical data, and treatment history.

4.2. Clinical evaluation results

The impact of AI-generated medical reports on ophthalmologists' treatment decisions was assessed across four categories: *Continue treatment* (Class 0), *Adjust treatment* (Class 1), *Change medication* (Class 2), and *Stop injections* (Class 3). As shown in Table 1, the most consistent benefit appeared in Class 0, where both evaluators achieved higher precision, recall, and F1-scores with AI support, reducing the risk of undertreatment. Class 1 contained too few cases for meaningful analysis, but results for Class 2 highlight how AI assistance can guide more confident medication adjustments: Evaluator 1 markedly improved precision (42% to 75%), while Evaluator 2 maintained comparable performance. For Class 3, recall was unchanged, yet the modest improvements in precision and F1-scores suggest AI may also contribute to safer treatment discontinuation decisions.

Beyond class-specific effects, AI support clearly enhanced overall decision quality. As presented in Table 2, accuracy improved from 0.42 to 0.61 for Evaluator 1 and from 0.42 to 0.52 for Evaluator 2. On average, accuracy across both evaluators increased from 0.42 (manual) to 0.56 (AI-assisted). The change ratios of 29% and 24% indicate that roughly one quarter to one third of treatment decisions differed when AI-generated reports were available, underscoring their tangible influence on clinical practice. Agreement between evaluators also strengthened, with Cohen's Kappa improving from 0.42 (manual) to 0.71 (AI-assisted). As detailed in Table 3, decision-making time was reduced by 43.8% for Evaluator 1 and 32.9% for Evaluator 2, corresponding to a mean reduction of 39%. Confidence levels remained stable across both evaluators. Together, these findings suggest that integrating AI-generated reports into routine workflows could reduce variability, improve consistency among specialists, and accelerate clinical decision-making without undermining trust.

Class	Evaluator	Manual			AI		
		Precision	Recall	F1	Precision	Recall	F1
0	1	0.50	0.50	0.50	0.60	0.90	0.72
	2	0.44	0.40	0.42	0.53	0.70	0.61
2	1	0.42	0.50	0.46	0.75	0.50	0.60
	2	0.50	0.66	0.57	0.50	0.50	0.50
3	1	1.00	0.33	0.50	1.00	0.33	0.50
	2	0.33	0.33	0.33	0.50	0.33	0.40

Table 1: Per-class comparison of manual and AI-assisted treatment decision metrics for both evaluators. Precision, recall, and F1-scores are reported for classes 0 (Continue), 2 (Change medication), and 3 (Stop treatment).

Evaluator	Manual Accuracy	AI Accuracy	Change Ratio
1	0.42	0.61	0.29
2	0.42	0.52	0.24

Table 2: Accuracy and change ratio for each evaluator with and without AI assistance. The change ratio indicates the proportion of treatment decisions that changed when comparing manual review with AI-assisted evaluation.

Evaluator	Time (min)		Confidence	
	Manual	AI	Manual	AI
1	1.94 ± 0.86	1.09 ± 0.34	3.27 ± 0.67	3.21 ± 1.62
2	1.55 ± 0.97	1.04 ± 0.24	4.78 ± 0.44	4.22 ± 0.45

Table 3: Comparison of decision-making time and confidence scores for each evaluator with and without AI assistance.

ISBN: 978-84-09-80259-3

The qualitative evaluation of the AI-generated medical reports yielded positive feedback across all dimensions. Clinicians rated their usefulness at 4.5/5, highlighting clinical relevance, support in treatment identification, and improved patient follow-up, with particular appreciation for the fluid progression plots, OCT charts, and segmentation analysis, although B-scan illustrations of major and minor fluids were seen as less informative. Usability was also valued, with a learning curve of 4.5/5 and clarity of 4/5, indicating that the reports are intuitive and require minimal training. Confidence in their accuracy, transparency, and ethical standards was uniformly strong, and clinicians expressed comfort integrating AI-generated information without compromising autonomy. Overall satisfaction and intention for future use were rated at 4.5/5, confirming that the reports are considered a reliable, user-friendly, and impactful tool for routine AMD patient management, with minor improvements in visualization suggested to further enhance usability.

4.3. Discussion

STEP-AMD demonstrates the feasibility of advancing beyond algorithmic prototypes and commercial tools limited to screening or isolated biomarker quantification. Building upon recent progress in OCT-based automation and remote monitoring, STEP-AMD integrates biomarker segmentation, treatment response prediction, and clinical data into longitudinal reports co-developed with clinicians. While research prototypes often perform well under controlled conditions, few undergo real-world validation. Our proof-of-concept indicates that AI-generated reports can enhance accuracy, agreement, and efficiency in AMD management, while being well accepted by clinicians.

Limitations include the computational cost of segmentation, which requires GPU acceleration for real-time use, and evaluation on a relatively small, imbalanced dataset. The system is currently deployed at the HGUGM for research and validation, operating on a secure virtual machine but not yet integrated with the electronic health record (EHR). Broader, prospective validation across larger cohorts and diverse imaging protocols is planned to confirm robustness and enable future EHR interoperability for routine clinical use.

5. Conclusions

STEP-AMD integrates AI-based OCT analysis, treatment prediction, and clinical data into structured reports for AMD follow-up. Deployed in a secure hospital environment for research and validation, it improved efficiency, interobserver agreement, and decision accuracy, while receiving positive clinician feedback.

Future work will focus on achieving EHR interoperability for automated data exchange and expanding the system with updated or complementary models, such as cube-based OCT segmentation and therapy-specific predictors. Extending validation to larger, prospectively collected cohorts and involving more clinicians will be essential to confirm robustness and support the translation of STEP-AMD into a practical tool for personalized AMD management.

Acknowledgments

The authors acknowledge the support of Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, under grants TED2021-131951B-I00 and PID2022-141493OB-I00 (10.13039/501100011033/MCIN/AEI/ERDF,UE), cofinanced by European Regional Development Fund (ERDF), 'A way of making Europe' and the Next Generation EU funds. AJBC is supported by a FPI grant from the Ministerio de Ciencia e Innovación - PREP2022-000162. The authors gratefully acknowledge the Universidad Politécnica de Madrid (www.upm.es) for providing computing resources on Magerit Supercomputer.

References

[1] P. Mitchell, G. Liew, B. Gopinath, and T.-Y. Wong, "Age-related macular degeneration," *The Lancet*, vol. 392, pp. 1147–1159, 09 2018.

[2] C. Metrangolo, S. Donati, M. Mazzola, L. Fontanel, W. Messina, G. D'alterio, M. Rubino, P. Radice, E. Premi, and C. Azzolini, "Oct biomarkers in neovascular age-related macular degeneration: A narrative review," *Journal of ophthalmology*, vol. 2021, p. 9994098, 07 2021.

[3] U. Schmidt-Erfurth, S. Klimscha, S. M. Waldstein, and H. Bogunović, "A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration," *Eye*, vol. 31, no. 1, pp. 26–44.

[4] Z. Mulyukov, P. Keane, J. Sahni, S. Liakopoulos, K. Hatz, D. Ting, R. Gallego-Pinazo, T. Aslam, C. M. G. Cheung, G. De Salvo, O. Semoun, G. Somfai, A. Stahl, B. Lujan, and D. Lorand, "Artificial intelligence-based disease activity monitoring to personalized neovascular age-related macular degeneration treatment: A feasibility study," *Ophthalmology Science*, vol. 4, p. 100565, 06 2024.

[5] T.-C. Yeh, A.-C. Luo, Y.-S. Deng, Y.-H. Lee, S.-J. Chen, P.-H. Chang, C.-J. Lin, M.-C. Tai, and Y.-B. Chou, "Prediction of treatment outcome in neovascular age-related macular degeneration using a novel convolutional neural network," *Scientific Reports*, vol. 12, 04 2022.

[6] M. Gallardo, M. Munk, T. Kurmann, S. De Zanet, A. Mosinska, I. Kutlutürk, M. Zinkernagel, S. Wolf, and R. Sznitman, "Machine learning can predict anti-vegf treatment demand in a treat-and-extend regimen for patients with namd, dme and rvo associated me," *Ophthalmology Retina*, vol. 5, 05 2021.

[7] C. Sastre, A. J. B. Carrero, J. T. Rodríguez, E. Santos, M. J. A. Hernández-Lastras, M. J. Ledesma-Carbayo, and J.-J. Gómez, "Design and implementation of an algorithm for the automated analysis of age-related macular degeneration biomarkers on optical coherence tomography."

[8] A. J. Beltrán-Carrero, J. Torresano-Rodríguez, E. Santos-Vicente, M. J. A. Hernández-Lastras, Á. Caballero-Sastre, M. J. Ledesma-Carbayo, and J. J. Gómez-Valverde, "Classification and prediction of age-related macular degeneration progression using oct images and multiple instance learning," in *International Challenge on Device-Independent Diabetic Macular Edema Onset Prediction*, ser. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2024, pp. 201–210.